



random, continuous noise signal  $n(k)$ , the noise component of the audio signal, wherein  $n(k)$  can include stationary and non-stationary noise components:

$$x(k) = s(k) + n(k) \quad (1)$$

5 A known method of removing or reducing random continuous noises of that kind, the so-called method of 'short time spectral attenuation' - referred to hereinafter for the sake of brevity as Short Time Spectral Attenuation (STSA) is shown in the block circuit diagram of Figure 1. Shown therein is the processing of an audio signal  $x(k)$  which is obtained as a  
10 sampling signal  $x(k)$  of the analog noisy audio signal  $x(t)$  at the sampling times  $k$ .

$X(m,l), S(m,l)$  and  $N(m,l)$  are the functions corresponding to the discrete signals  $x(k), s(k)$ , and  $n(k)$ , for example in the frequency domain, wherein  $m$  denotes the discrete frequency. Alternatively however  $m$  can be  
15 another parameter which permits equivalent description of the discrete time signals  $x(k), s(k)$ , and  $n(k)$ .  $l$  is the discrete time of the respective signal block being considered, with conventional block-wise signal processing. Therefore the following correspondingly applies in the frequency domain:

$$20 \quad X(m,l) = S(m,l) + N(m,l) \quad (2)$$

In this known method the discrete audio signal  $x(k)$  is transformed in a first step by means of a discrete Fourier transform into the frequency domain, block 1, so that the discrete frequency domain representation  $X(m,l)$  is the result. In the illustrated state of the art, that discrete spectral  
25 representation affords a single and thus stationary estimate  $\hat{\Phi}_{NN}(m)$  of the discrete auto-noise power density  $\Phi_{NN}(m)$  by a known estimation process, block 2, which for example involves:

- (3a) an estimate of the auto-noise power density within (approximately) useful signal-free passages of the noisy signal, or  
30 (3b) a so-called direct estimate.

The estimated discrete auto-noise power density  $\hat{\Phi}_{NN}(m)$  comes from a discrete, randomly continuously noisy audio signal in accordance with the process referred to in (3a) by evaluation of approximately audio signal-free passages of the noisy signal, in which as an approximation the following  
5 applies:

$$x(k) \approx n(k), \text{ as } s(k) \approx 0. \quad (3)$$

Making use of the linearity of the Fourier transform there is within those portions in which  $s(k) \approx 0$ , an estimate of the discrete auto-noise power density, in accordance with the following:

$$\hat{\Phi}_{NN}(m) = \Phi_{XX}(m) \quad (4)$$

Therein  $\Phi_{XX}(m)$  denotes the auto-noise power density of the noisy audio signal.

The alternative process (3b) referred to as 'direct estimate' was presented in 'Steven L Gay, Jacob Benesty: *Acoustic Signal Processing for*  
15 *Telecommunication*; Kluwer International Series in Engineering and Computer Science; Chapter 9; Eric J Diethorn: *Subband Noise Reduction Methods for Speech Enhancement*, March 2000, ISBN 0-7923-7814-8' and is based on limitedly tracking the power density of the noisy signal.

In that known process, based on the estimate of the auto-noise  
20 power density  $\hat{\Phi}_{NN}(m)$  and the discrete frequency domain representation  $X(m, l)$  of the discrete audio signal  $x(k)$ , there is determined a suitable filter function  $H_G(m, l)$ , see block 3, in which the delivered signal approximates as accurately as possible to the non-noisy audio signal  $s(k)$ . In this connection various calculation procedures are known for obtaining the filter  
25 function  $H_G(m, l)$ , for example:

(6a) the approach in accordance with Wiener, in which the mean quadratic error between useful signal and estimate is used as the approximation criterion, or

(6b) the approach relating to amplitude subtraction, or

30 (6c) the approach relating to power subtraction

which are described in 'S F Boll; *Suppression of acoustic noise in speech using spectral subtraction*; IEEE Trans Acoust, Speech & Signal Process.; ASSP-27, pages 113-120; 1979', and also in the textbook by P Vary, U Heute & W Hess '*Digitale Sprachsignalverarbeitung*', Teubner Verlag, Stuttgart 1998, ISBN 3-519-06165-1, pages 380-390.

Determining an estimate  $\hat{s}(k)$  of the discrete non-noisy useful component  $s(k)$  involves effecting filtering of the discrete audio signal  $x(k)$  with the previously determined filter function. That can be implemented either in the time domain by convolution of the discrete noisy signal  $x(k)$  with the discrete pulse response of the filter function  $h_G(k)$ :

$$\hat{s}(k) = h_G(k) * x(k), \quad (5)$$

wherein  $*$  represents the convolution operator or as shown in Figure 1 in the frequency domain by multiplication of the discrete transfer function  $H_G(m, l)$  with the discrete spectral representation  $X(m, l)$  of the discrete noisy audio signal  $x(k, l)$ , see block 4:

$$\hat{S}(m, l) = H_G(m, l) \cdot X(m, l). \quad (6)$$

Using the discrete estimate  $\hat{S}(m, l)$  determined in that way, the corresponding representation  $\hat{s}(k)$  is obtained therefrom in the time domain by the inverse discrete Fourier transform, see block 5, so that the noise-freed signal can be converted, possibly by means of a digital-analog converter, into an analog, noise-freed signal.

A disadvantage of that known method is that the operation of filtering the noisy audio signal causes noise to be again introduced into the noise-freed signal, which occurs due to the filtering operation and results in unwanted so-called 'musical tones'.

In addition, 'M Berouti, R Schwartz & J Makhoul: *Enhancement of speech corrupted by acoustic noise*; in Proc. IEEE ICASSP; page 208-211; Washington DC; 1979' discloses a further method which is described hereinafter with reference to the block circuit diagram of Figure 2 and which corresponds in terms of its basic principle to the method shown in Figure 1. That known method operates in the following manner:

Taking a single and thus stationary estimate of the auto-noise power density  $\hat{\Phi}_{NN}(m)$ , block 2, and the discrete signal representation  $X(m,l)$  at the output of the block 1 of the discrete audio signal  $x(k)$ , the filter function  $H_G(m,l)$  is ascertained therefrom, block 3. Prior to the actual filtering of the noisy signal, block 4, the filter function  $H_G(m,l)$  is limited to a constant, freely selected minimum value  $\gamma_{SF}(m)$  - also referred to as the 'spectral bottom' -, that is to say a maximum noise reduction, block 6. That therefore affords for the filtering operation a new discrete filter function  $H_G(m,l,\gamma_{SF}(m))$ , for which the following applies:

$$H_G(m,l,\gamma_{SF}(m)) = \begin{cases} H_G(m,l) & \text{for } H_G(m,l) > \gamma_{SF}(m) \\ \gamma_{SF}(m) & \text{other} \end{cases} \quad (7)$$

That limited filter function means on the one hand that no freedom from noise but only a reduction in interference is possible, while on the other hand the occurrence of so-called musical tones is markedly reduced.

The discrete, noise-reduced signal spectrum  $\hat{S}(m,l)$  obtained by the filtering operation, block 4, is then transferred back into the time domain as in the method shown in Figure 1 by inverse discrete Fourier transform, block 5.

Both known methods are found to suffer from the disadvantage that they can only be used for the removal or reduction of random, continuous, stationary and possibly random, continuous, slowly non-stationary noise. Changes in respect of time of the statistical properties of the discrete noise  $n(k)$  cannot be detected or can be detected only in the case of very slow changes. If however the superimposed interference involves for example a non-stationary noise, that affords an error-inflicted estimate of the auto-noise power density. That results in defective determination of the filter function and thus a noise reduction which either adversely affects the

actual non-noisy signal  $s(k)$  and/or only insufficiently reduces the noise signal  $n(k)$ .

When using a one-off and thus stationary estimate of the auto-noise power density within useful signal-free portions, there is a defective auto-  
 5 noise power density as a random continuously disturbed audio signal generally does not have sufficiently many useful signal-free portions which permit continuous updating of the estimate of the auto-noise power. This means that the estimate value ascertained cannot take account of the changes in respect of time of the statistical properties of the noise.  
 10 Admittedly, with the above-discussed and known 'direct estimate' the auto-noise power density is continuously updated, but the estimate is defective in respect of the non-stationary noise component, as is shown by the considerations in that respect in 'J Meyer, K U Simmer and K D Kammeyer: *Comparison of One- and Two-Channel Noise-Estimation Techniques; Proc*  
 15 *5th International Workshop on Acoustic Echo and Noise Control (IWAENC-97)*, Vol 1, pages 17-20, London, UK 11-12th September 1997'.

US patent No 5 852 567 discloses a further method of reducing random continuous noise. Based on a time-frequency transform the endeavour with that method is to improve the signal-noise ratio and the  
 20 characteristics of the non-stationary useful signal. As in the methods described hereinbefore, this method is also found to suffer from the disadvantage that, in accordance with its development aim, it can also only be used for reducing random continuous stationary noise but not for reducing random continuous non-stationary noise.

25 Therefore the object of the invention is to provide a method and an apparatus for producing random continuous non-stationary noise, with the aim of reducing the non-stationary noise component in the audio signal in relation to the stationary noise component thereof.

That object is attained by a method as set forth in claim 1. In  
 30 addition that object is attained by an apparatus as set forth in claim 11.

The advantages of the method according to the invention and the apparatus according to the invention are that a representation of the noisy audio signal is processed in such a way that the changes in respect of time



Figure 4 is a block circuit diagram of a first embodiment of the method according to the invention,

Figure 5 is a block circuit diagram of a second embodiment of the method according to the invention,

5 Figure 6 is a block circuit diagram of a third embodiment of the method according to the invention,

Figures 7a through 7c show the typical configuration in respect of time of the noise component a) of a noisy audio signal, b) of the audio signal processed in accordance with the state of the art, and c) of the audio  
10 signal processed with the method according to the invention,

Figure 8 is a representation by way of example of the mode of operation of the method shown in Figure 2,

Figure 9 is a diagrammatic view of the mode of operation of an embodiment of the known method when using an estimate of the currently  
15 contained noise signal component which describes the change in respect of time of the noise for determining the filter function  $H_G^{dyn}(m,l)$  and the restriction thereof by means of a restriction function  $\gamma_{SF}(m)$  which is constant in respect of time, and

Figure 10 is a representation by way of example of the mode of  
20 operation of an embodiment of the method according to the invention.

Figures 3 and 4 show a diagrammatic block circuit diagram of a first embodiment of the method according to the invention. In accordance with the block circuit diagram shown in Figure 3, the procedure involves determining from a discrete noisy audio signal  $x(k)$  by a suitable transform,  
25 for example a transform of the signal  $x(k)$  into the frequency domain, an associated representation  $X(m,l)$  of that audio signal, block 1. The variable  $l$  describes in this connection the current observation time. That representation is processed in a processing unit 2. The processing of that representation, in accordance with the method of the invention, affords the  
30 processed new representation  $\hat{S}(m,l)$  of the audio signal which is characterised by a reduction in the changes in respect of time of the statistical properties of the contained noise component. Finally then by

suitable reverse transformation the discrete signal configuration  $\hat{s}(k)$  is obtained, which describes the discrete configuration in respect of time of the noise-reduced audio signal as a function of the discrete sampling times.

As shown in Figure 4 a suitable filter function  $H_G^{dyn}(m,l)$  is determined  
 5 from the representation of the noisy audio signal  $X_2(m,l)$  - which for example is afforded by a suitable imaging procedure from the representation  $X(m,l)$  and which represents the signal  $x(k)$  transformed from the time domain into the frequency domain - see block 5, and the representation  $\hat{N}(m,l)$  which represents an estimate of the current  
 10 properties of the noise signal component in the frequency domain, in known manner, utilising the estimate  $\hat{N}(m,l)$  of the noise component of the audio signal. In addition, utilising the estimate  $\hat{N}(m,l)$  of the noise component of the audio signal, the filter function  $H_G^{dyn}(m,l)$  ascertained in that way is restricted dynamically, that is to say in dependence on time,  
 15 see blocks 4 and 6. The superscript *dyn* characterises a filter function which is obtained by incorporating the current properties of the non-stationary noise component of the audio signal.

In a further processing step the representation  $X(m,l)$  of the noisy audio signal  $x(k)$  is filtered with the restricted filter function, see block 7,  
 20 thus affording a processed discrete signal  $\hat{S}(m,l)$ . That representation  $\hat{S}(m,l)$ , by means of suitable reverse transform, affords a discrete signal configuration  $\hat{s}(k)$  which corresponds to the discrete configuration in respect of time of the noisy audio signal  $x(k)$ , but is characterised by a smaller change in respect of time of the statistical properties of the  
 25 contained noise.

Figure 5 shows the block circuit diagram relating to the implementation of a second embodiment of the method according to the invention. The procedure involves ascertaining from the discrete noisy audio signal  $x(k)$  at the respective observation time  $l$ , for example by a  
 30 Fourier transform, a suitable representation  $X(m,l)$  of that audio signal,

see block 1. Obtained therefrom is an estimate  $\hat{N}(m,l)$  of the non-stationary random and continuous noise component  $n(k)$  which is superimposed on the non-noisy discrete audio signal  $s(k)$ , see block 4, which describes the current statistical properties of the non-stationary noise. Using the estimate  $\hat{N}(m,l)$ , a suitable filter function  $H_G^{dyn}(m,l)$ , see block 8, which in contrast to the known methods takes account of the non-stationary nature of the interference component, is ascertained utilising the representation of the noisy signal  $X(m,l)$  - which is possibly additionally changed by a suitable imaging procedure (not shown). In the following step that filter function  $H_G^{dyn}(m,l)$  is restricted to a minimum value  $\gamma_{SF}(m,l)$ , see block 9. That limit - also referred to as the restriction function - is not constant but is determined dynamically in dependence on a direct or indirect representation of the interference noise:

$$\gamma_{SF}(m,l) = f(\hat{N}(m,l)) \quad (8)$$

A representation of the noisy audio signal  $x(k)$  can particularly preferably additionally also be used for the calculation of  $\gamma_{SF}(m,l)$ . The following then applies:

$$\gamma_{SF}(m,l) = f(\hat{N}(m,l), X(m,l)) \quad (9)$$

The following then applies for the filter function  $H_b$  which is restricted in that way:

$$H_b = H_G^{dyn}(m,l, \gamma_{SF}(m,l)) = \begin{cases} H_G^{dyn}(m,l) & \text{for } H_G^{dyn}(m,l) > \gamma_{SF}(m,l) \\ \gamma_{SF}(m,l) & \text{other} \end{cases} \quad (10)$$

A suitable linking - for example a multiplication procedure - of a representation  $X(m,l)$  of the noisy audio signal  $s(k)$  with the previously ascertained restricted filter function  $H_b = H_G^{dyn}(m,l, \gamma_{SF}(m,l))$  then supplies a discrete signal  $\hat{S}(m,l)$  from which it is possible to derive, by reverse transform corresponding to the transform, a discrete signal sequence  $\hat{s}(k)$

which corresponds to the noisy audio signal  $x(k)$ , but is characterised by a smaller change in respect of time of the statistical properties of the contained noise, see block 6.

Figure 6 shows a block circuit diagram of a third embodiment of the method according to the invention which serves for the reduction of a random continuous non-stationary noise in an audio signal which is adversely affected by amplitude-modulated noise interference with constant spectral coloration. The discrete spectrum  $X(m,l)$  of the noisy audio signal is obtained at the observation time  $l$ , see block 10, from the discrete noisy audio signal  $x(k)$  by a fast Fourier transform (FFT).  $X(m,l)$  is also referred to as the representation form of the noisy audio signal. On the basis of that discrete spectrum  $X(m,l)$  an estimate is effected in respect of the auto-noise power density  $\hat{\Phi}_{NN}(m,l)$ , applicable at the observation time  $l$ , which is a measurement in respect of the noise component  $n(k)$  in the noisy audio signal  $x(k)$ . That estimation procedure is effected in two steps:

- in a first step, an estimate value  $\hat{\Phi}_{NN}(m)$  of the stationary auto-noise power density is ascertained by one of the known estimation procedures, the power density describing the spectral coloration but not the configuration in respect of time of the interference noise, block 22;
- then a second step involves ascertaining a parameter which characterises the non-stationary nature of the noise, block 24. For that purpose, there is determined from the estimated auto-noise power density  $\hat{\Phi}_{NN}(m)$  and the spectrum  $X(m,l)$  of the noisy audio signal a time-variant modulation factor  $\alpha(m,l)$  which describes the amplitude modulation of the noise, for example:

$$\alpha(m,l) = \frac{\min(|X(m,l)|^2)}{\min(\hat{\Phi}_{NN}(m))} \quad (11)$$

Multiplication of the estimated stationary auto-noise power density  $\hat{\Phi}_{NN}(m)$  by that modulation factor then affords the wanted estimate value  $\hat{\Phi}_{NN}(m,l)$  of the actual auto-noise power density  $\Phi_{NN}(m,l)$ , block 26:

$$\hat{\Phi}_{NN}(m,l) = \alpha(m,l) \cdot \hat{\Phi}_{NN}(m). \quad (12)$$

5 On the basis thereof, with the incorporation of the current discrete Fourier transforms  $X(m,l)$  of the noisy audio signal  $x(k)$  the procedure involves determining a filter function  $H_G^{dyn}(m,l)$  for the current observation time  $l$  by means of a suitable approach, for example by means of the known approach in accordance with Wiener, block 30.

10 The filter function  $H_G^{dyn}(m,l)$  is restricted hereafter by means of a restriction function  $\gamma_{SF}(m,l)$  dynamically adapted to the properties of the noise, in terms of its amplitude, which for example from the previously calculated modulation factor  $\alpha(m,l)$ , in accordance with:

$$\gamma_{SF}(m,l) \sim (\alpha(m,l))^\beta \quad (13)$$

15 with  $-5 < \beta < +5$ ;  $\beta = -1/2$  is particularly preferred, behaves in proportional manner, block 40.

Then, the dynamically restricted filter function  $H_b$  can be determined by means of the restriction function obtained in that way, in accordance with equation (10), block 40.

20 Then, in a further step, the discrete Fourier transforms of the noisy signal  $X(m,l)$  is multiplied by the previously ascertained restricted filter function  $H_b$ , see block 50. Finally, by inverse fast Fourier transform (IFFT) it is possible to determine from the resulting estimate  $\hat{S}(m,l)$  a signal  $\hat{s}(k)$ , block 60, which corresponds to the noisy audio signal by reduced modulation of the noise, namely a smaller change in respect of time of the statistical properties of the contained noise, and is characterised by a noise reduction which is dependent on the restriction function  $\gamma_{SF}(m,l)$ .

Figure 7a shows the variation in respect of time of a noise component  $n(k)$  which is superimposed on any discrete non-noisy useful

component  $s(k)$ . If a discrete randomly, continuously and non-stationarily  
noisy audio signal  $x(k) = s(k) + n(k)$  which is composed in that way is  
processed by means of a known method as referred to in the preamble to  
the description, that affords a noise component which is shown in Figure  
5 7b. If in comparison the audio signal  $x(k)$  which has non-stationary noise is  
processed with the method according to the invention, then, after the  
processing operation, that gives the resulting noise component shown in  
Figure 7c, which is of a stationary character which is uniform in relation to  
time; the typical non-stationarity of the signal, which is present in Figures  
10 7a and 7b, has been successfully eliminated as shown in Figure 7c.

To explain the mode of operation of the method according to the  
invention, the basic starting point adopted hereinafter will be an audio  
signal  $x(k)$  which is processed in block-wise manner and whose  
representation  $X(m, l)$  corresponds to the square of the block-wise Fourier  
15 transform. The audio signal  $x(k)$  is to comprise a non-stationary noise  $n(k)$   
or  $N(m, l)$  and is not to contain any useful signal  $s(k)$ . Accordingly the  
following applies for the discrete frequency  $m_i$  (with  $i = 1, 2, 3, \dots$ ) and the  
discrete times  $l$ , which are associated with the individual signal blocks:

$$X(m_i, l) = N(m_i, l) \quad (14)$$

20

By way of example, the associated illustrations, Figures 8a, 9a and  
10a, reproduce the configuration in respect of time  $N(m_i, l)$  for a discrete  
frequency  $m_i$ .

When using the known method with restricted STSA, taking the  
25 stationary estimate of the auto-noise power density  $\hat{N}(m_i)$ , shown in  
broken line in Figure 8a, and the noise signal, a filter function  $H_G$  is  
calculated by means of a suitable method (for example in accordance with  
Wiener), Figure 8b. In the regions in which the real noise representation  
 $N(m_i, l)$  falls below the stationary estimate  $\hat{N}(m_i)$ , the filter function  
30  $H_G(m_i, l)$  assumes a value close to zero and the noise interference is

approximately completely suppressed at those times  $l$ . In contrast, for those times  $l$  in which the representation of the real noise power density  $N(m_i, l)$  is greater than the estimate, the filter function  $H_G(m_i, l)$  assumes a value of close to one as a part of the current noise signal is interpreted as a useful signal.

If that filter function is limited in accordance with the STSA method to a constant lower limit  $\gamma_{SF}(m_i)$  which is therefore invariable in respect of time, that gives a configuration in respect of time as shown in Figure 8c. If the filter function  $H_G(m_i, l, \gamma_{SF}(m_i))$  produced in that way is applied to the interference noise signal, that again gives as the output signal a non-stationary residual noise, see Figure 8d.

Figure 9 represents the diagrammatic mode of operation of the method illustrated in Figure 8, in which however the representation, which was estimated on a one-off basis and is thus stationary, of the auto-noise power density  $\hat{N}(m_i)$ , is replaced by a dynamic estimate of the auto-noise power density  $N(m_i, l)$ , that is to say an estimate which describes the variations in respect of time of the noise. As the filter function  $H_G(m_i, l)$  for example by adopting the Wiener approach, there is obtained a function which is fixed by a constant restriction function  $\gamma_{SF}(m_i)$  in accordance with equation (7) at a lower limit which is invariable in respect of time, see Figure 9c. If the filter signal is subjected to filtering with the restricted filter function  $H_G(m_i, l, \gamma_{SF}(m_i))$ , then the processed signal, as shown in Figure 9b, contains a residual noise whose amplitude is markedly reduced in comparison with the amplitude shown in Figure 8d, but in which case the non-stationarity of the noise signal is not removed.

If the method described with reference to Figures 9a through 9d is supplemented by a further step, that gives the method according to the invention as shown in Figure 10. If the filter function  $H_G(m_i, l)$ , as shown in Figure 9b, is restricted by means of a restriction function  $\gamma_{SF}(m_i, l)$  which is variable in respect of time, for example in accordance with equation (13), it is possible to achieve a residual noise in the output signal, which is almost

or completely stationary, and which therefore no longer includes the non-stationarity in respect of time of the signal  $n(k)$ . The filter function  $H_G^{dyn}(m, l)$  is determined from the estimate  $\hat{N}(m, l)$  which describes the change in respect of time of the noise, Figure 10a, and from the noisy signal  $X(m, l)$ , see Figure 10b. That function is restricted by a restriction function  $\gamma_{sf}(m, l)$  which is variable in respect of time, in accordance with equation (10), so that this affords the dynamically restricted filter function  $H_b = H_G^{dyn}(m, l, \gamma_{sf}(m, l))$  in accordance with equations (10) and (13), see Figure 10c. Filtering of the input signal with that filter function now results in a processed signal which only still contains a stationary residual noise, see Figure 10d.